# DETECTION OF THE DERMIS/EPIDERMIS BOUNDARY IN REFLECTANCE CONFOCAL IMAGES USING MULTI-SCALE CLASSIFIER WITH ADAPTIVE TEXTURE FEATURES

*Sila Kurugol*[1]*, Jennifer Dy* [1]*, Milind Rajadhyaksha* [2] *and Dana H. Brooks* [1]

(1) ECE Dept., Northeastern University, 360 Huntington Ave., Boston, MA
(2) Dermatology Service, Memorial Sloan-Kettering Cancer Cent., 1275 York Ave., New York, NY

## ABSTRACT

Reflectance confocal microscopy is an emerging modality for dermatology applications, especially in-situ and bedside detection of skin cancers. Work to date has concentrated on hardware development and validation by clinicians in comparison with standard histological staining. As this technology gains acceptance, the development of automated processing methods becomes more important. We concentrate here on detection of the dominant internal feature of the skin, the epidermis/dermis boundary, a complex corrugated 3-dimensional layer marked by optically subtle changes and features. We adopt a machine learning approach to this segmentation problem, using a hierarchical multi-scale classifier with sophisticated on-line feature selection, to minimize the required expert labeling and maximize the range of potential features in the face of high inter- and intra-subject variability and low optical contrast. Initial results indicate the ability of our approach to recover the complex 3-D boundary surface.

***Index Terms—*** confocal microscopy, image segmentation, classification

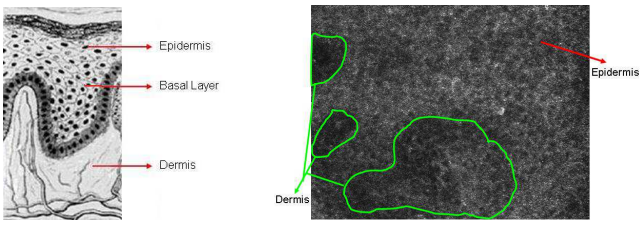
## 1. INTRODUCTION

Skin cancer is the most common form of cancer in the U.S. Clinical screening and diagnosis is based on biopsy followed by histology. Biopsies are invasive, painful, destroy the site and leave a scar. Confocal reflectance microscopy (CRM) is a new modality for optical imaging of the skin in vivo and shows potential for non-invasive clinical screening and diagnosis without biopsy. Nuclear and cellular detail is imaged non-invasively with optical sectioning and $\mu$m-level resolution that is comparable to that of histology [1]. CRM is moving toward clinical application in dermatology for detection of malignancies such as melanoma and basal cell carcinomas [2, 3]. However, interpretation of these images is dependent on highly trained clinicians, and since they appear quite different from conventional stained pathology sections, significant clinician re-training is required. Thus, any automated processing, even if only partial, which could aid diagnosis would be a significant contribution to the utility of skin CRM. However, as we describe below, interpretation of CRM images presents significant difficulty for automation.

As an initial target for automatic interpretation of CRM images, in this paper we investigate semi-automatic localization of the irregular three-dimensional junction between the superficial epidermis and the underlying dermis in acquired CRM image z-stacks. This is a meaningful target application because it is both important and presents difficulties such as low contrast and high variability which are typical of CRM images of the skin. The localization of the Dermis/Epidermis junction (DEJ) is important for skin cancer since pre-malignancies such as melanoma and basal cell carcinoma start at the DEJ and then advance to cancer laterally and/or invade deeper into dermis. The latter, deeper invasion, is more serious than the initial superficially lateral spread. Thus the DEJ is clinically important to evaluate. This is also true for other tissue besides skin. Delineation of the DEJ is visually difficult in grayscale monochromatic confocal reflectance images (i.e., without any exogenous specific stains), as the contrast between the deeper epidermis (basal cell layer) and the underlying dermis is relatively low. Consequently, visual and subjective detection of the junction is difficult for dermatologists, as noted in ongoing clinical studies [2, 3]. In contrast, specific stains are used to provide cellular and nuclear-specific contrast in pathology methods. Egg carton-like structure of the DEJ is illustrated in Fig. 1 (a). The epidermis itself has a depth-dependent layered structure composed of cells with different morphology at each layer, while the upper dermis mostly contains collagen fibers and blood vessels.

Factors which make the DEJ difficult to detect in CRM images include not only the lack of reliable optical contrast at this boundary but also the tremendous variability among subjects and even, to some extent, within sites of the same subject. One critical factor is skin pigmentation, which depends on melanin, also located at this boundary. Melanin provides strong reflectance contrast where it is abundant, but in fair

Support for the work of MR provided in part by R01EB002715 from NIH/NIBIB's Image Guided Interventions program (Prog. Direct. Dr. John Haller). The work of the SK, JD, and DHB was supported in part by the Center for Communications and Digital Signal Processing Research at Northeastern University and SK and DHB were also supported by the NIH/NCRR Center for Integrative Biomedical Computing (CIBC), P41-RR12553-07, JD was also partly supported by NSF IIS-0347532




(a) (b)

**Fig. 1**. (a) Skin: Epidermis, basal layer, junction and dermis (b) CRM slice showing dermis and epidermis regions.

skin this contrast is almost absent, as illustrated in a CRM image from a subject with fair skin in Fig. 1 (b). Our goal is a method which can work across a range of skin pigmentations, and therefore we have designed a processing scheme which contains considerable flexibility to include a variety of potential features.

Additional difficulties result from the nature of confocal imaging in skin. In particular, we need to image to depths of over $100\ \mu m$, and the hills and valleys of the boundary itself can extend over a depth of $50\ \mu m$, which requires repeatedly increasing the illumination power as we image at greater depth to maintain sufficient dynamic range. Moreover optical aberrations increase at depth due to scattering, so the resulting images lose resolution, and of course axial resolution is considerably worse than transverse resolution.

The approach we present below starts with some preprocessing for normalization. We then localize the slices which contain DEJ regions via a rapid 1D criterion. To then segment these slices into epidermis and dermis, we require that the user label several regions of dermis and epidermis on a single image. Using those labeled regions, a superset of texture features is extracted. Our method then automatically selects the features with higher discriminant power and applies a hierarchical classification scheme. In particular, we repeatedly subdivide the images into tiles up to a preset minimum size, attempting after each subdivision to classify the tiles as dermis or epidermis. Unclassifiable tiles are subject to further subdivision. In addition we incorporate the constraint that the transition from epidermis to dermis is unidirectional with increasing depth. We show initial results on data from two subjects with the two fairest, *i.e.* most challenging, skin types.

## 2. METHODS

A CRM stack from a single site consists of about 60 image slices, where each slice is on the order of $1000 \times 1000$ pixels, with a pixel resolution of $0.5\mu m$. The upper slices are composed of only epidermis regions and the lower slices are composed of only dermis. As a result the $z$-stack image collection for a single site is on the order of 60 MB. Given the low contrast of the desired boundary, and large variability among individuals and acquisitions, we require a large set of features and yet a computationally efficient processing scheme.

Our algorithm is semi-automatic; typically the expert will label a small initial set of dermis and epidermis regions in one slice which contains a significant number of regions of both types; if necessary these regions can come from two near-by slices. The training data used to initiate feature selection comes from these labeled regions. To augment this training data, additional epidermis data is automatically collected from regions in a small number of slices directly above the expert-selected epidermis regions, and similarly additional dermis data from a small number of slices in regions directly below the expert-selected dermis regions. The rest of the processing proceeds in an automated fashion using this training data.

In order to trade-off feature reliability vs. detection resolution, we use a multi-scale approach. We start by applying a coarse-level classification along the optical axis ($z$-direction), which trims off the pure epidermis and pure dermis slices. We, then, apply a multi-scale classification within each of the remaining mixed dermis/epidermis slices using an adaptive feature selection methodology.

Before beginning this process, we perform pre-processing steps which involve manually masking out uninteresting features such as wrinkles and sweat glands and normalization to compensate for both loss of illumination intensity with depth and staggered increases in applied laser power. The normalization is based on standard histogram matching [4].

### 2.1. Coarse-Level Classification in the $z$-Direction

The goal in this step is to detect two boundaries: the lower boundary of pure epidermis slices and the upper boundary of pure dermis slices. We extract features (the same features described in Sec. 2.3) from each image slice and construct a data matrix from these features for all slices in the stack. We apply PCA [5] to the data matrix and represent each image slice by the projection of the feature vector from a slice onto the first principal component. We then model these 1-D data across the $z$-stack by a 1-D autoregressive (AR) model of order two. We calculate the fit error between the AR predictions and the true value. The two highest peaks in this error vector indicates the potential boundary locations.

### 2.2. Multi-Scale Classification Within Each Slice

First we describe the multi-scale hierarchy we use, followed by a description of the classification scheme applied at each hierarchy level. At the coarsest level of the hierarchy, we classify tiles of size $s_{max} \times s_{max}$ (here, $100 \times 100$) pixels. At this level, some tiles contain both epidermis and dermis regions, so we include a third class, mixed dermis/epidermis, along with dermis and epidermis classes. After classification at the

coarsest level, we split each tile into four smaller tiles of size $s_{max}/2 \times s_{max}/2$. At each scale $i$, we use a specific classifier trained for that scale to classify the tiles at scale $i$ into one of these three classes. If a tile is classified as mixed, we split it further into four smaller tiles. If the tile is classified as a pure dermis or pure epidermis, we check for consistency with the next level down the finer resolution tree to be sure that it is indeed pure and not mixed. If the finer level classification is consistent with the previous level, we do not split further; otherwise we continue splitting. We continue this process until we reach our finest level with tile size $s_{min} \times s_{min}$. Here, we set $s_{min} = 25$ ($12.5\ \mu m$) because, beyond this scale, we lose the characteristic features containing information about the intrinsic structures in the tissue.

We train a different classifier for each scale. We investigated two classification approaches. The first was an ensemble of 1-D support vector machines (SVM) [6] based on each feature alone, where the final classification was obtained by majority vote. The second method was SVM using all the selected features. (We describe the features and how features were selected in the next subsection.) We trained our SVM classifier on the two labeled classes, dermis and epidermis. We modified SVM to include the mixed class by assigning points located between the support vectors to that class. The ensemble of single feature SVM method performed better on this data; due to space constraints, we report only those results here.

### 2.3. Automatic Feature Selection

We used a number of texture features to capture varying texture structures to enhance the classifier's ability to perform well across different skin pigmentations. We used as features statistics, namely mean, variance, skewness, and kurtosis, energy calculated from an entropy filter output, gray level co-occurrence matrix contrast, energy, correlation and homogeneity, features obtained from a stationary wavelet decomposition up to level 3 (energy, entropy, variance) [7] and features like the average power spectrum in a relevant frequency band [8]. Note that the feature set is computed for each scale.

At each scale we customize this large set of features to the current image stack by automatically selecting a subset of non-redundant features that best discriminates the training data. Since we automatically select features on-line, we need an efficient selection procedure. We tested sequential forward search using SVM as the classifier, with the selection criterion as the margin between the separating hyper-planes. This wrapper approach gave high accuracy, but this process is slow when the feature set is large. Thus, we also tested a faster filter method, the correlation based filter approach [9] which is based on two steps: 1) Find the features that are relevant to the class, 2) find the subset of the relevant features that is least redundant to the other relevant features. For the first step, we use Fisher's class separation distance ($D_x(c_1, c_2)$ (Eq. 1)) for feature $x$ between class 1 and class 2, as the measure of relevance, instead of using correlation between features and class, because Fisher's criterion not only finds the features relevant to the class but also keeps the more separable ones. The filter method searches for features one at a time. The first step selects the features with $D_x$ larger then a certain threshold; the second step keeps the features which are not highly correlated with the other features by evaluating the correlation coefficient between each pair $(x, y)$.

$$D_x(c_1, c_2) = \frac{|\mu_{c_1} - \mu_{c_2}|}{\sqrt{\sigma_{c_1}^2 + \sigma_{c_2}^2}} \quad (1)$$

## 3. RESULTS

We report results for data from two subjects with fair skin types—skin types I and II. These are the hardest skin types in which to find the DEJ as noted earlier. We used ten $100 \times 100$ manually selected training regions, 5 dermis and 5 epidermis, augmented as described above by data from regions in 10 slices above and 10 slices below the labeled epidermis and dermis regions respectively. Thresholds for class separation and correlation were set off-line. The transition regions for these two data sets were found by the $z$-direction classification, as described, to be between slices 6 and 44 (out of 75 total) for the first data set, and between 23 and 46 (out of 60) for the second data set.

Fig. 2 shows six representative segmented slices, three from data set 1 (left column) and three from data set 2 (right column). The regions classified as epidermis are outlined with green lines and shaded with a light-colored overlay. Fig. 2(b) is the expert-segmented slice for this data set, and the white lines show the hand-drawn boundaries of the dermis regions in this slice. The dark regions in all images are wrinkles.

To show the 3-D boundary that results from our segmentation, we visualize the detected surface in Fig. 3 for both data sets. This surface is generated by finding the slice number for each tile where the transition from epidermis to dermis takes place. We then interpolate the 3-D surface to create a smoother boundary. In addition, we truncate the graph below for better visibility of the boundary (which otherwise would be dominated by the dark wrinkles).

We note that the detection results in each slice correspond well to visual inspection (as evidenced by the agreement with the expert-labeled slice in Fig. 2). The 3-D boundary shown in Fig. 3 has the expected "egg-carton" shape.

## 4. CONCLUSION AND FUTURE WORK

A hierarchical multi-scale classification algorithm with automatically selected adapted features are used to localize the DEJ. After roughly locating the junction layer location in the

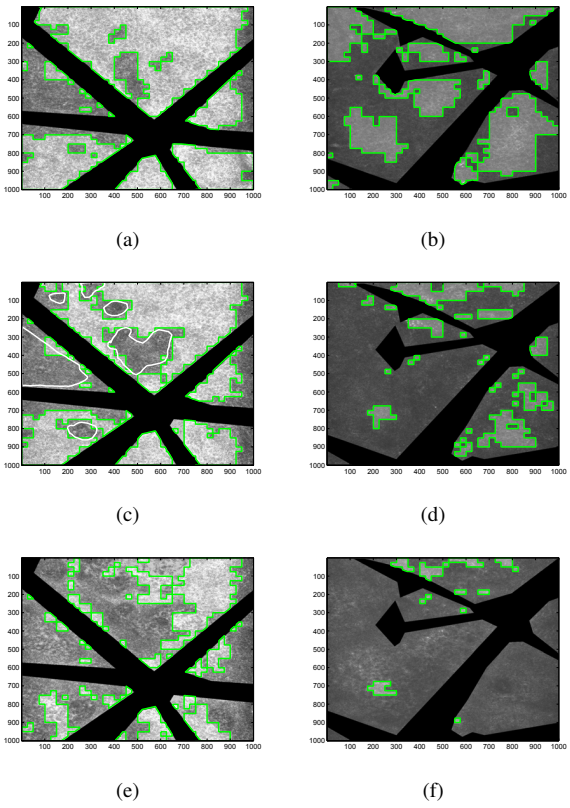


**Fig. 2**. Example 2-D images of classification results for both data sets. Data set 1 is in the left column and data set 2 in the right column. The slices shown are numbers 20 (a), 25 (c), and 30 (e) for data set 1 and numbers 32 (b), 45 (d), 50 (e) for data set 2.

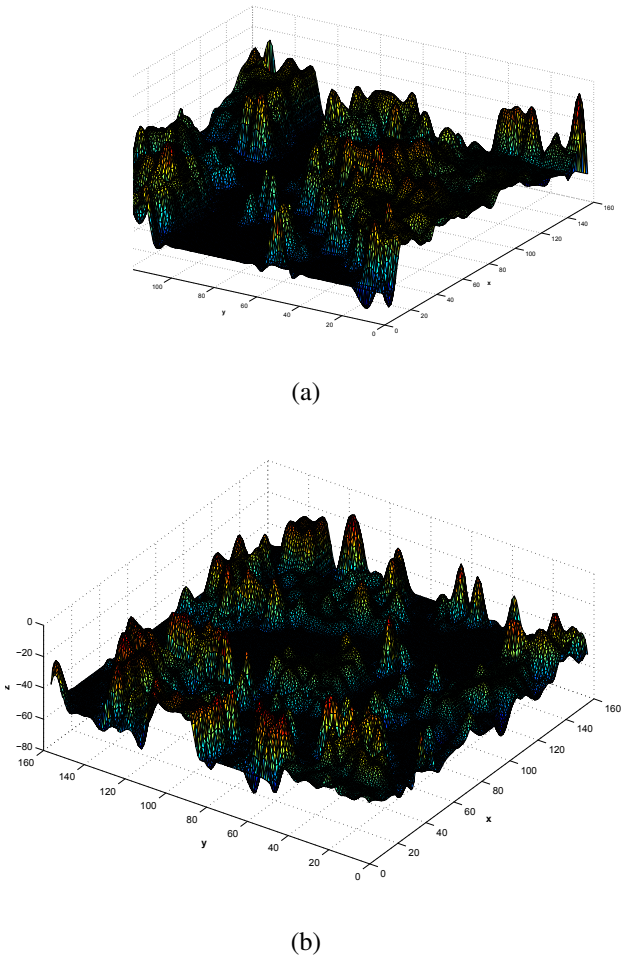


**Fig. 3**. Surface plot of detected dermo-epidermal junction in 3-D for (a) data set 1 and (b) data set 2

$z$-direction, junction layer slices are classified in 2-D by splitting them into tiles at each scale. Due to the low contrast between structures, unlike histology images where stains are used, in CRM images it is very difficult to find reliable features, hence choosing the right features and adding more prior knowledge to the method is required for adequate results. A possible approach for improvement is to use more specific features at the cellular scale, since cellular structure in the epidermis vs. blurry collagen distributed in the dermis is an important visual clue for discriminating these regions.

**Acknowledgments** Special thanks to Jocelyn Lieb, M.D. for helping us to collect data and Allan Halpern, M.D. for his advices.

## 5. REFERENCES

[1] M. Rajadhyaksha, S. Gonzalez, J.M. Zavislan, R. R. Anderson, and R. H. Webb, "In vivo confocal scanning laser microscopy of human skin II," *J. Invest. Dermatol*, vol. 113, pp. 293–303, 1999.

[2] G. Pellacani et al, "The impact of in vivo reflectance confocal microscopy for the diagnostic accuracy of melanoma and equivocal melanocytic lesions.," *J. Invest. Dermatol*, vol. 127, pp. 2759–65, 2007.

[3] S. Nori et. al, "Sensitivity and specificity of reflectance-mode confocal microscopy for in vivo diagnosis of basal cell carcinoma: A multicenter study," *J. American Academy of Dermatol*, vol. 51, pp. 923–930, 2004.

[4] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Prentice Hall, 2002.

[5] I.T. Jolliffe, *Princip. Component Analysis*, Prentice Hall, 1986.

[6] Vapnik V.N., *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.

[7] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," 1995, pp. 281–300, in Lecture Notes in Statistics.

[8] M. Tuceryan and A.K. Jain, "Texture analysis," *in Handbook of Pattern Recog. and Comp. Vis.*, pp. 235–276, 1993.

[9] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Research*, vol. 5, pp. 1205–1224, 2004.